ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΔΗΜΟΣΙΑΣ ΔΙΟΙΚΗΣΗΣ & ΑΥΤΟΔΙΟΙΚΗΣΗΣ
ΙΝΣΤΙΤΟΥΤΟ ΕΠΙΜΟΡΦΩΣΗΣ

Στατιστική Συμπερασματολογία με Στατιστικά Πακέτα

Παρουσίαση Εκπαιδευτή

Μαθησιακό Αντικείμενο:

Περιγραφική Στατιστική ΙΙ
Μέτρα Συμπύκνωσης των πληροφοριών

Εκπαιδευτικοί στόχοι

Με την υλοποίηση του μαθησιακού αντικειμένου, ο καθένας από τους συμμετέχοντες θα μπορεί:

- Να απαριθμεί τα στατιστικά μέτρα και να δίνει τους ορισμούς τους.
- Να υπολογίζει και να ερμηνεύει τα στατιστικά μέτρα τάσης ή θέσης.
- Να υπολογίζει και να ερμηνεύει τα στατιστικά μέτρα διασποράς:
- Να υπολογίζει και να ερμηνεύει τα στατιστικά μέτρα ασυμμετρίας και κύρτωσης.
- Να αξιολογεί τη δομή μονομεταβλητών πληθυσμών με περιγραφικά στατιστικά εργαλεία ανάλυσης.
- Να εφαρμόζει τα παραπάνω χρησιμοποιώντας το στατιστικό πακέτο SPSS.

Παράμετροι τάσης (1)

Αριθμητικός μέσος (arithmetic mean)
(αναφέρεται απλά ως μέσος ή ο μέσος των δεδομένων)

Δειγματικός μέσος → στο **δείγμα**

→ $\overline{X}$: Ο μέσος των παρατηρήσεων $X_1, X_2, \ldots, X_n$

Μέση τιμή → στο **γενικό πληθυσμό**

→ **μ** : Η μέση τιμή των τιμών $x_1, x_2, \ldots, x_N$ των N μονάδων του πληθυσμού

n: αριθμός των παρατηρήσεων ενός **δείγματος** και
N: αριθμός των μονάδων ενός **πληθυσμού**

Παράμετροι τάσης (2)

Ο μαθηματικός τύπος του **αριθμητικού μέσου** είναι ο παρακάτω:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \text{και} \quad \mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Ο μαθηματικός τύπος του **σταθμισμένου αριθμητικού μέσου** είναι ο παρακάτω:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_k x_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum f_i x_i}{\sum f_i}$$

όπου $x_1, x_2, \ldots, x_k$ οι τιμές της μεταβλητής X και $f_1, f_2, \ldots, f_k$ οι συχνότητες των τιμών της μεταβλητής X.

Παράμετροι τάσης (3)

Γεωμετρικός Μέσος

Ο γεωμετρικός μέσος μιας σειράς παρατηρήσεων $x_1, x_2, \ldots, x_N$ ορίζεται ως η νιοστή ρίζα του γινομένου τους.

$$G = \sqrt[N]{x_1 \cdot x_2 \cdot \ldots \cdot x_N}$$

Σε περίπτωση συνεχούς μεταβλητής, λαμβάνονται υπ' όψη οι συχνότητες ως σταθμίσεις και ο τύπος γίνεται:

$$G = \sqrt[\sum f_i]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_N^{f_N}} \quad \text{ή} \quad \lambda o \gamma G = \frac{\sum_{i=1}^{N} f_i \lambda o \gamma x_i}{\sum f_i}$$

Παράμετροι θέσεις (1)

Διάμεσος (median)

Η **διάμεσος (median)** ενός συνόλου μετρήσεων είναι η τιμή εκείνη με την ιδιότητα ότι το πολύ 50% των μετρήσεων είναι μικρότερες από την τιμή αυτή και το πολύ 50% των μετρήσεων είναι μεγαλύτερες από την τιμή αυτή.

Παράμετροι θέσεις (2)

Με μία παραλλαγή του τύπου της διαμέσου μπορούμε να εξάγουμε σημαντικές πληροφορίες από τα στατιστικά μας δεδομένα όπως:

- πάνω από ποια τιμή βρίσκεται το **75%** των παρατηρήσεων του πληθυσμού (***πρώτο τεταρτημόριο***)
- πάνω από ποια τιμή βρίσκεται το **25%** (***τρίτο τεταρτημόριο***)
- κάτω από ποια τιμή βρίσκεται το **25%** (**πρώτο τεταρτημόριο**)
- κάτω από ποια τιμή βρίσκεται το **75%** (**τρίτο τεταρτημόριο**).

Οι τιμές αυτές είναι δηλαδή **οι θέσεις** του πρώτου και του τρίτου τετάρτου των παρατηρήσεων του πληθυσμού.

Παράμετροι θέσεις (3)

Επικρατούσα Τιμή (mode)

Η τιμή εκείνη των δεδομένων που έχει τη μεγαλύτερη συχνότητα εμφάνισης.

- ✓ Μια μόνο επικρατούσα τιμή ονομάζεται **μονοκόρυφη (unimodal).**
- ✓ Δύο επικρατούσες τιμές ονομάζεται **δικόρυφη (bimodal).**
- ✓ Η επικρατούσα τιμή χρησιμοποιείται σε δεδομένα ποιοτικά.

Πλεονεκτήματα – Μειονεκτήματα Μέσης Τιμής

Πλεονεκτήματα	**Μειονεκτήματα**
Για τον υπολογισμό της χρησιμοποιούνται όλες οι τιμές.	*Επηρεάζεται πολύ από ακραίες τιμές.*
Είναι μοναδική για κάθε σύνολο δεδομένων.	*Μπορεί να μην αντιστοιχεί σε δυνατή τιμή της μεταβλητής.*
Είναι εύκολα κατανοητή.	*Δεν υπολογίζεται για ποιοτικά δεδομένα.*
Ο υπολογισμός της είναι σχετικά εύκολος.	*Είναι δύσκολος ο υπολογισμός της σε ομαδοποιημένα δεδομένα* *με ανοικτές τις ακραίες κλάσεις.*
Αξιοποιείται στη στατιστική συμπερασματολογία.	

Πλεονεκτήματα – Μειονεκτήματα Διαμέσου

Πλεονεκτήματα	**Μειονεκτήματα**
Είναι εύκολα κατανοητή.	*Δε χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της.*
Δεν επηρεάζεται από ακραίες τιμές.	*Είναι δύσκολη η αξιοποίησή της στη στατιστική συμπερασματολογία.*
Υπολογίζεται και στην περίπτωση που οι ακραίες κλάσεις είναι ανοικτές.	*Δεν υπολογίζεται για κατηγορικά δεδομένα.*
Ο υπολογισμός της είναι απλός.	*Για τον υπολογισμό της μπορεί να χρειαστεί παρεμβολή.*
Είναι μοναδική σε κάθε σύνολο δεδομένων.	

Πλεονεκτήματα – Μειονεκτήματα Επικρατούσας Τιμής

Πλεονεκτήματα	**Μειονεκτήματα**
Υπολογίζεται εύκολα.	*Δε χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της.*
Είναι εύκολα κατανοητή.	*Στη στατιστική συμπερασματολογία έχει περιορισμένη σημασία.*
Υπολογίζεται και από ελλιπή δεδομένα.	*Δεν ορίζεται πάντα μονοσήμαντα. Δηλαδή, μπορεί να υπάρχουν περισσότερες από μία ή και καθόλου.*
Δεν επηρεάζεται από ακραίες τιμές.	
Υπολογίζεται και για ποιοτικά δεδομένα.	

Παράμετροι Διασποράς (1)

Έκταση ή **Εύρος (Range)**

Ως **έκταση ή εύρος (range)** ορίζουμε τη διαφορά μεταξύ της μεγαλύτερης και της μικρότερης τιμής ενός συνόλου δεδομένων.

Δηλαδή

$$R = X_{max} - X_{min}$$

όπου

X_{max}: η μέγιστη τιμή του συνόλου των δεδομένων και

X_{min}: η ελάχιστη τιμή του συνόλου αυτού.

Παράμετροι Διασποράς (2)

Μέση Απόκλιση

Η *μέση απόκλιση* ορίζεται ως ο μέσος αριθμητικός όλων των απολύτων διαφορών των τιμών μιας μεταβλητής από το μέσο αριθμητικό (μ) της μεταβλητής.

$$M.A. = \frac{\sum |x_i - \bar{x}|}{N}$$

Όπου $|X_i - \bar{X}|$ συμβολίζει την απόλυτη τιμή της διαφοράς μεταξύ της τιμής X_i των δεδομένων και του μέσου των δεδομένων αυτών $\bar{X}$.

Παράμετροι Διασποράς (3)

Διακύμανση ή Διασπορά (Variance)

Ορίζουμε ως **διακύμανση** ή **διασπορά (variance)** ενός πληθυσμού N τιμών $x_1, x_2, \dots, x_N$ με μέση τιμή μ, **τη μέση τετραγωνική απόκλιση των N παρατηρήσεων από τη μέση τιμή μ του πληθυσμού.**

Για τη διακύμανση χρησιμοποιείται ο συμβολισμός σ^2.

Επομένως έχουμε

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Παράμετροι Διασποράς (4)

Τυπική Απόκλιση (standard deviation)

Ορίζουμε ως **τυπική απόκλιση (standard deviation)** τη θετική τετραγωνική ρίζα της διακύμανσης.

Θα έχουμε όσον αφορά τον **πληθυσμό** $\sigma = \sqrt{\sigma^2}$

Και όσον αφορά το **δείγμα** $S = \sqrt{S^2} \ \text{και} \ S^* = \sqrt{S^{*2}}$

Ούτε η διακύμανση ούτε η τυπική απόκλιση μπορούν να έχουν **αρνητικές τιμές.**

Παράμετροι Διασποράς (5)

Συντελεστής Μεταβλητικότητας

Όταν προσπαθούμε να συγκρίνουμε μεταβλητές ή κατανομές μεταβλητών οι οποίες εκφράζονται **σε διαφορετικές μονάδες** μέτρησης η τυπική απόκλιση σαν μέτρο διασποράς παρουσιάζει αδυναμίες.

Χρησιμοποιούμε το **συντελεστή μεταβλητικότητας**, όπου εκφράζεται ως ο λόγος της τυπικής απόκλισης μιας κατανομής προς τον αριθμητικό μέσο.

$$CV = \frac{\sigma}{\bar{x}} \cdot 100$$

Παράμετροι Ασυμμετρίας (1)

Για τη μέτρηση του βαθμού ασυμμετρίας μιας κατανομής έχουν προταθεί κυρίως από τους Bowley και Pearson οι παρακάτω συντελεστές ασυμμετρίας:

- *Συντελεστής* ***Bowley***

$$S_k(B) = \frac{(Q_3 - M_e) - (M_e - Q_1)}{(Q_3 - M_e) + (M_e - Q_1)} = \frac{Q_3 + Q_1 - 2 \cdot M_e}{Q_3 - Q_1}$$

όπου M_e: Η διάμεση τιμή.

Παίρνει τιμές μεταξύ της αρνητικής και της θετικής μονάδας.

Παράμετροι Ασυμμετρίας (2)

- Μέτρο του **Pearson**.

μέσος - επικρατούσα τιμή = 3(μέσος - διάμεσος)

Έχει το *μειονέκτημα* ότι προϋποθέτει **γνώση της επικρατούσας τιμής** η οποία δεν είναι πάντα εύκολο να προσδιοριστεί, κυρίως στην περίπτωση των ομαδοποιημένων δεδομένων.

Λόγω της συγκεκριμένης δυσκολίας και δεδομένου ότι για μία ελαφρά ασύμμετρη κατανομή ισχύει κατά προσέγγιση η σχέση:

μέσος - επικρατούσα τιμή = 3(μέσος - διάμεσος)

Συχνά χρησιμοποιείται ο συντελεστής

$$\frac{3(\text{μέσος} - \text{διάμεσος})}{\text{τυπική απόκλιση}}$$

Παράμετροι Κύρτωσης (1)

Η *κύρτωση* μιας κατανομής μετράει το βαθμό συγκεντρώσεως των τιμών της μεταβλητής στην περιοχή των άκρων και του μέσου αριθμητικού.

Για τον προσδιορισμό του βαθμού κύρτωσης χρησιμοποιούμε το **συντελεστή σχετικής κύρτωσης** που δίνεται από τον τύπο:

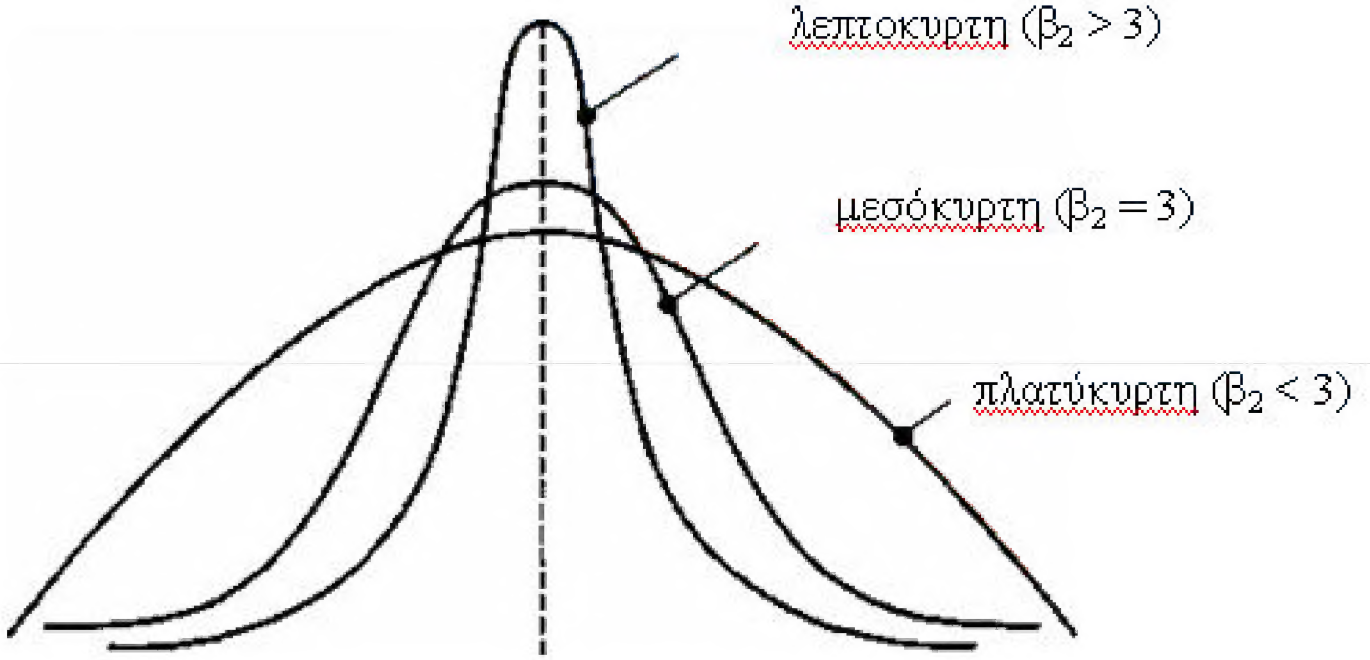
$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{(\sigma^2)^2}$$

Όπου β_2: ο συντελεστής σχετικής κύρτωσης

μ_2: τιμή της δεύτερης κεντρικής ροπής

μ_4: τιμή της τέταρτης κεντρικής ροπής

Παράμετροι Κύρτωσης (2)



Διάγραμμα 1: Καμπύλες Συχνοτήτων τριών κατανομών